

# Professor SVD

VON CLEVE MOLER

Gene Golub ist Computerwissenschaftler an der Stanford University. Er hat mehr als jeder andere dazu beigetragen, die Singulärwertzerlegung zu einem der mächtigsten und verbreitetsten Werkzeuge der modernen Linearen Algebra zu machen.



Gene Golubs Nummernschild. Fotografie von Professor P.M. Kroonenberg von der Universität Leiden.

Die Singulärwertzerlegung (SVD) ist eine relative neue Entwicklung. Laut Pete Stewart, dem Autor des 1993 erschienen Artikels "On the Early History of the Singular Value Decomposition", wurde der Begriff *valeurs singulières* zum ersten Mal um 1910 von Emile Picard im Zusammenhang mit Integralgleichungen verwendet. Picard verwendete hier das Adjektiv „singulär“

in der Bedeutung „außergewöhnlich“ oder „außerordentlich“. Ein Bezug zu singulären Matrizen bestand zu diesem Zeitpunkt noch nicht.

Als ich in den frühen sechziger Jahren Doktorand war, wurde die SVD immer noch als ein ziemlich obskures theoretisches Konzept angesehen. In einem Buch, das George Forsythe und ich 1964 gemeinsam schrieben, wurde die SVD als eine relativ schlechte

Methode beschrieben, die Norm und die Konditionszahl einer Matrix zu kennzeichnen. Einen praktischen Weg, SVDs zu berechnen, gab es noch nicht. Gene Golub und W. Kahan veröffentlichten 1965 den ersten Algorithmus, der das änderte. Die heute verwendete Form der SVD basiert auf einer 1970 von Gene Golub und Christian Reinsch veröffentlichten Vari-

ante dieses Algorithmus. Als 1980 die erste MATLAB®-Version erschien, war die SVD eines ihrer Highlights. Lassen Sie uns ein Beispiel in Form einer 2x2-Matrix erzeugen, indem wir den Vorgang umkehren und eine Matrix aus ihrer SVD berechnen. Wir setzen  $\sigma_1 = 2$ ,  $\sigma_2 = 1/2$ ,  $\theta = \pi/6$  und  $\phi = \pi/4$ . Es seien

$$U = \begin{pmatrix} -\cos \theta & \sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}$$

$$V = \begin{pmatrix} -\cos \phi & \sin \phi \\ \sin \phi & \cos \phi \end{pmatrix}$$

Die Matrizen  $U$  und  $V$  beschreiben *Drehungen* um die Winkel  $\theta$  und  $\phi$ , gefolgt von *Spiegelungen* in der ersten Dimension. Die Matrix  $\Sigma$  ist eine *diagonale* Skalentransformation. Wenn man nun die Matrix  $A$  nach

$$A = U\Sigma V^T$$

berechnet, erhält man das Ergebnis

$$A = \begin{pmatrix} 1.4015 & -1.0480 \\ .4009 & 1.0133 \end{pmatrix}$$

Es besagt, dass die Matrix  $A$  erzeugt wird durch eine 45°-Drehung und eine Spiegelung, gefolgt von unabhängigen Skalierungen der beiden Koordinatenrichtungen mit dem Faktor 2 beziehungsweise  $\frac{1}{2}$  und schließlich durch eine 30°-Drehung, gefolgt von einer weiteren Spiegelung.

Die MATLAB-Funktion `eigshow` erzeugt ein Diagramm, das die Singulärwertzerlegung einer 2x2-Matrix demonstriert.

Die Singulärwertzerlegung (SVD) ist eine Faktorisierung von Matrizen mit einer Vielzahl interessanter Anwendungen.

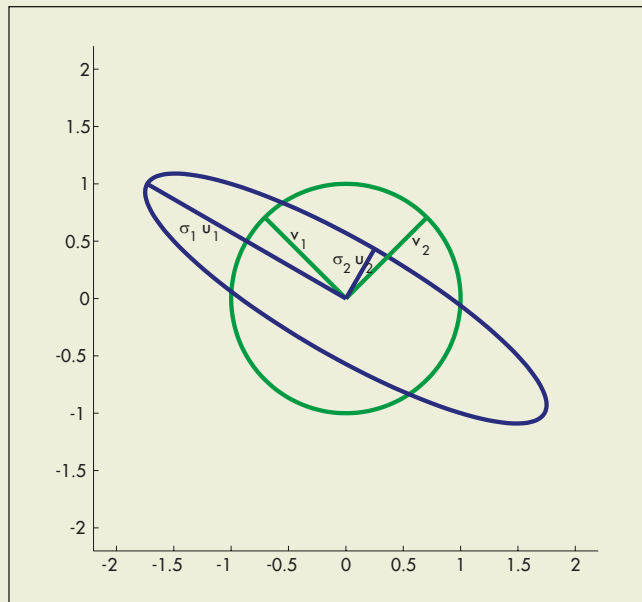


Abb. 1: Die von eigshow erzeugte SVD-Darstellung.

Wir geben dazu folgende Anweisungen ein:

```
A = [1.4015 -1.0480;
     -0.4009  1.0133]
eigshow(A)
```

Wenn man nun den Knopf „SVD“ klickt und die Maus etwas bewegt, erscheint Abbildung 1, wenn auch mit etwas anderer Beschriftung.

Der grüne Kreis ist der Einheitskreis in der Ebene. Die blaue Ellipse stellt das Bild des Kreises nach Transformation durch die Matrix  $A$  dar. Die grünen Vektoren  $v_1$  und  $v_2$ , die für die Spalten von  $V$  stehen, und die blauen Vektoren  $u_1$  und  $u_2$ , die für die Spalten von  $U$  stehen, sind zwei verschiedene orthogonale Basen für den zweidimensionalen Raum.

Die Spalten von  $V$  sind um  $45^\circ$  gegen die Achsen der Abbildung gedreht, während die Spalten von  $U$ , die die große und kleine Halbachse der Ellipse bilden, um  $30^\circ$  gedreht sind. Die Matrix  $A$  transformiert  $v_1$  in  $\sigma_1 u_1$  und  $v_2$  in  $\sigma_2 u_2$ .

Lassen Sie uns nun zu Matrizen der Ordnung  $m$  mal  $n$  übergehen. Eine der wichtigsten Eigenschaften der SVD ist ihre Anwendbarkeit auf orthogonale Matrizen. Eine reelle Matrix  $U$  ist orthogonal oder hat orthonormale Spalten, wenn gilt:

$$U^T U = I$$

Die Spalten von  $U$  stehen also paarweise senkrecht aufeinander und haben Einheitslänge. Geometrisch betrachtet sind Transformationen mit orthogonalen Matrizen eine Verallgemeinerung von Drehungen und Spiegelungen; alle Längen und Winkel bleiben dabei erhalten. Rechnerisch sind orthogonale Matrizen attraktiv, weil sie Rundungsfehler und andere Fehler nicht vergrößern. Jede reelle Matrix  $A$ , sogar eine nicht-quadratische Matrix, kann als Produkt dreier Matrizen geschrieben werden.

$$A = U \Sigma V^T$$

Die Matrix  $U$  ist orthogonal und hat genau so viele Spalten wie  $A$ . Die Matrix  $\Sigma$  hat die gleiche Größe wie  $A$ , besitzt aber nur auf ihrer Diagonale Elemente, die ungleich Null sind. Die Diagonalelemente von  $\Sigma$  sind die *Singulärwerte* und die Spalten von  $U$  und  $V$  sind die rechten und linken *Singulärvektoren*.

In der abstrakten Sprache der Linearen Algebra stellt eine Matrix eine lineare Transformation eines Vektorraums, den *Definitionsbereich*, in einen anderen Vektorraum, den *Wertebereich*, dar. Die SVD besagt also, dass es für jede lineare Transformation möglich ist, eine orthonormale Basis für den Definitionsbereich und eine (möglicherweise andere) orthonormale Basis für den Wertebereich zu wählen. Die Transforma-

tion wird damit unabhängig von allen Skalierungen oder Dehnungen in den Koordinatenrichtungen.

Der *Rang* einer Matrix ist die Zahl linear unabhängiger Zeilen, die wiederum identisch mit der Zahl linear unabhängiger Spalten ist. Der Rang einer Diagonalmatrix ist also offensichtlich gleich der Zahl der Diagonalelemente, die ungleich Null sind. Orthogonale Transformationen haben auf die Zahl linear unabhängiger Zeilen oder Spalten keinen Einfluss. Der Rang jeder Matrix ist damit immer gleich der Zahl aller Singulärwerte, die ungleich Null sind. Wir geben nun in MATLAB die Anweisung

`type rank`

ein, um zu untersuchen, wie man geschickt Toleranzen wählen kann und die nicht vernachlässigbar kleinen Singulärwerte findet. In traditionellen Kursen in Linearer Algebra wird meist noch die Zeilenreduzierte Stufenform (Reduced Row Echelon Form, RREF) besonders hervorgehoben. Wenn man aber mit ungenauen Daten und mit in-exakter Arithmetik arbeiten muss, ist diese Methode wenig zuverlässig. Die SVD ist eine Art moderner Ersatz für die RREF. Eine quadratische Matrix ist *nichtsingulär*, wenn, und nur wenn, alle Diagonalelemente darin ungleich Null sind. Der numerisch zuverlässigste Weg zur Bestimmung, ob Matrizen singulär sind, besteht darin, ihre Singulärwerte zu testen. Das geht erheblich einfacher, als ihre Determinanten zu berechnen, die fürchterliche Skalierungseigenschaften haben.

Mit Hilfe der Singulärwertzerlegung wird das lineare Gleichungssystem

$$Ax = b$$

zu

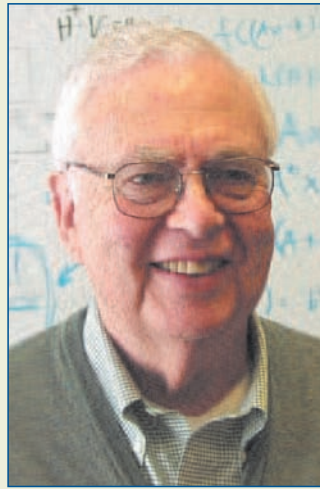
$$U \Sigma V^T x = b$$

Die Lösung dafür ist

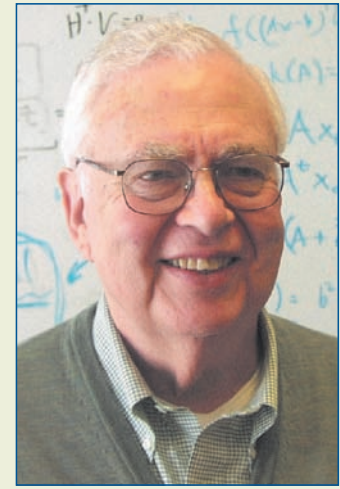
$$x = V \Sigma^{-1} U^T b$$



12



50



120

Abb. 2: Näherungen des Ranges 12, 50 und 120 des Fotos von Gene Golub mit dem Rang 598.

Man multipliziert mit einer orthogonalen Matrix, teilt durch die Singulärwerte und multipliziert dann mit einer anderen orthogonalen Matrix. Der Rechenaufwand ist zwar höher als beim Gauß-Algorithmus, dafür sind aber die numerischen Eigenschaften dieses Verfahrens exzellent. Man kann abschätzen, ob die Singulärwerte vernachlässigbar klein sind und für den Fall, dass das zutrifft, das entsprechende singuläre System analysieren.

$E_k$  sei das Vektorprodukt des k-ten linken und rechten Singulärvektors, also

$$E_k = u_k v_k^T$$

Dann kann A als eine Summe aus Matrizen vom Rang 1 ausgedrückt werden

$$A = \sum_{k=1}^n \sigma_k E_k$$

Wenn man nun die Singulärwerte in aufsteigender Reihenfolge ordnet,  $\sigma_1 > \sigma_2 > \dots > \sigma_n$ , und die Summe nach r Termen abschneidet, ist das Ergebnis eine *Näherung* der Ausgangsmatrix vom Rang r. Der Fehler dieser Näherung hängt von der Größenordnung der vernachlässigten Singulärwerte ab. Wenn man auf diese Weise mit einer Datenmatrix verfährt, die durch Subtraktion des Mittelwertes jeder Spalte von der gesamten Spalte zentriert wurde, nennt man diesen Vorgang *Hauptkomponentenanalyse*

(Principal Component Analysis, PCA). Die rechten Singulärvektoren,  $v_k$ , sind die *Komponenten*, die skalierten linken Singulärvektoren,  $\sigma_k u_k$ , sind die so genannten *Scores*. PCAs werden meistens mit Hilfe der Eigenwerte und Eigenvektoren der Kovarianzmatrix  $AA^T$  beschrieben. Manchmal hat der SVD-Ansatz aber bessere numerische Eigenschaften.

SVDs und Matrizen-Approximationen werden häufig am Beispiel der Kompression eines Bildes demonstriert. Für unser Beispiel verwenden wir Gene Golubs Foto von seiner Webseite (Abb. 2). Das Bild hat 897 mal 598 Pixel. Wir schreiben die Werte für die Rot- Grün- und Blauanteile einzeln untereinander, wodurch wir eine 2691-mal-598-Matrix erhalten. Wir führen dann eine einzige SVD durch. Nachdem wir so eine Näherung niedrigen Ranges durchgeführt haben, setzen wir die neuen RGB-Werte in das Bild ein. Wenn wir mit einem Rang von 12 arbeiten, stimmen bereits die Farben und man kann Gene erkennen. Das geht am besten, wenn man die Augen ein wenig zusammenkneift und das Gehirn das Bild rekonstruieren lässt. Bei einem Rang von 50 kann man bereits die Formeln auf der Tafel hinter Gene lesen. Wenn wir einen Rang von 120 wählen, gibt es kaum noch einen Unterschied zum maximalen Rang von 598. (Für eine echte Bildkompression eignet sich diese Methode allerdings weniger. Um ehrlich zu sein, meine Freunde aus

der Bildverarbeitung bezeichnen dies als „Bildverschlechterung“.)

Das Thema Eigenwerte wurde bis hierher praktisch noch gar nicht angeschnitten. Ich wollte nämlich zeigen, dass es möglich ist, über Singulärwerte zu reden, ohne dabei Eigenwerte zu erwähnen – obwohl beide natürlich eng mit einander verwandt sind. Wenn A quadratisch, symmetrisch und positiv definit ist, sind ihre Singulärwerte sogar mit den Eigenwerten identisch und die linken und rechten Singulärvektoren sind mit einander und mit den Eigenvektoren der Matrix identisch. Allgemeiner ausgedrückt sind die Singulärwerte von A die Quadratwurzeln der Eigenwerte von  $A^T A$  oder  $AA^T$ . Singulärwerte sind dann wichtig, wenn man eine Matrix als Transformation eines Raums in einen anderen Raum mit möglicherweise anderen Dimensionen betrachtet. Eigenwerte dagegen sind wichtig, wenn man eine Matrix als Transformation eines Raum in sich selbst ansieht, beispielsweise bei gewöhnlichen linearen Gleichungen.

Google findet über 3.000.000 Webseiten, die den Begriff „Singular Value Decomposition“ enthalten und fast 200.000 Seiten mit „SVD MATLAB“. Einige dieser Seiten kannte ich schon, bevor ich diese Kolumne zu schreiben begann. Ein paar andere interessante Seiten habe ich erst beim Surfen entdeckt.

Professor SVD hat all das (und vieles mehr) möglich gemacht. Vielen Dank, Gene. ◀

# Einige Suchergebnisse für den Begriff "Singular Value Decomposition"

- Die Wikipedia-Seiten zur SVD und PCA sind sehr gut und enthalten eine Reihe nützlicher Links, wenn auch nicht zueinander.

[en.wikipedia.org/wiki/Singular\\_value\\_decomposition](http://en.wikipedia.org/wiki/Singular_value_decomposition)

[en.wikipedia.org/wiki/Principal\\_component\\_analysis](http://en.wikipedia.org/wiki/Principal_component_analysis)

- Rasmus Bro, Professor der Königlichen Tierärztlichen und Landwirtschaftlichen Universität in Dänemark, und Barry Wise, Leiter der Eigenvektor-Forschung in Wenatchee, Washington, beschäftigen sich beide mit der Chemometrie mit Hilfe von SVDs und PCAs. Ein Beispiel zeigt die Analyse des Absorptionsspektrums von Wasser zur Identifizierung flussaufwärts gelegener Verschmutzungsquellen.

[www.models.kvl.dk/users/rasmus](http://www.models.kvl.dk/users/rasmus)

[www.eigenvector.com](http://www.eigenvector.com)

- Tammy Kolda und Brett Bader von den Sandia National Labs in Livermore, Kalifornien, haben die Tensor Toolbox für MATLAB entwickelt, mit der die PCA auf mehrdimensionale Datensätze verallgemeinert werden kann.

[csmlr.ca.sandia.gov/~tgkolda/TensorToolbox](http://csmlr.ca.sandia.gov/~tgkolda/TensorToolbox)

- 2003 hat Lawrence Sirovich von der Mount Sinai School of Medicine in den Proceedings of the U.S. National Academy of Sciences den Artikel "A pattern Analysis of the second Rehnquist U.S. Supreme Court" veröffentlicht. Sein Artikel war Ausgangspunkt für Beiträge in der New York Times und der Washington Post, weil er ein unpolitisches, phänomenologisches Modell von Gerichtsentscheidungen aufstellt. Zwischen 1994 und 2002 hörte das Gericht 468 Fälle. Da es neun Richter gibt, von denen jeder in jedem Fall eine Mehrheits- oder Minderheitsposition einnimmt, bilden die Daten eine 468-mal-9-Matrix von +1s und -1s. Wären alle Entscheidungen per Münzwurf gefallen, hätte die Matrix ziemlich sicher den Rang 9 gehabt. Sirovich fand aber, dass der dritte Singulärwert um eine Größenordnung kleiner ist als der erste. Die Matrix hat also in guter Näherung den Rang 2. Anders ausgedrückt befinden sich die meisten Entscheidungen des Gerichts näherungsweise in einem zweidimensionalen Unterraum sämtlicher möglichen Urteile.

[www.pnas.org/cgi/reprint/100/13/7432](http://www.pnas.org/cgi/reprint/100/13/7432)

- Bei der semantikgestützten Ähnlichkeitssuche sucht man nach Dokumenten, indem man die SVD auf Matrizen anwendet, die Wahrscheinlichkeiten des Auftauchens von Begriffen in einem Artikel enthalten. Sollte beispielsweise eine Suche nach „Singulärwert“ auch nach „Eigenwert“ suchen? Michael Berry, Zlato Drmac und Liz Jessup haben dazu 1999 den Artikel "Matrices, Vector Spaces, and Information Retrieval" in der SIAM Review veröffentlicht.

[epubs.siam.org/SIREV/volume-41/art\\_34703.html](http://epubs.siam.org/SIREV/volume-41/art_34703.html)

- Der erste Google-Treffer zu "protein svd" ist "Protein Substate Modeling and Identification Using the SVD", von Tod Romo von der Rice University. Die Seite stellt die Anwendung der SVD auf die Analyse der Struktur und der Bewegung von Proteinen vor und enthält einige wundervolle Grafiken.

[bioc.rice.edu/~tromo/Sprez/toc.html](http://bioc.rice.edu/~tromo/Sprez/toc.html)

- Die Biophysiker Micheal Wall, Andreas Rechsteiner und Luis Rocha (Los Alamos) geben einen guten Online-Überblick über die SVD und PCA am Beispiel ihrer Anwendungen auf die Genexpressions-Analyse.

[public.lanl.gov/mewall/kluwer2002.html](http://public.lanl.gov/mewall/kluwer2002.html)

- „Darstellung periodischer menschlicher Bewegungen mit Hilfe der Funktionsanalyse“, (2005), von Dirk Ormoneit, Michael Black, Trevor Hastie und Hedvig Kjellstrom, beschreibt Methoden, bei denen Fourierzerlegungen und PCAs zur Analyse und Modellierung der Motion-Capture-Daten von Tätigkeiten wie dem Laufen eingesetzt werden.

[www.csc.kth.se/~hedvig/publications/ivc\\_05.pdf](http://www.csc.kth.se/~hedvig/publications/ivc_05.pdf)

- Ein ähnliches Thema hat der Artikel „Zerlegung biologischer Bewegungen: ein Gerüst für die Analyse und Synthese menschlicher Gangmuster“, (2002), von Nicholas Troje. Trojes Arbeit ist auch die Basis für eine "Eigenwalker"-Demo.

[www.journalofvision.org/2/5/2](http://www.journalofvision.org/2/5/2)

[www.mathworks.com/moler/ncm/walker.m](http://www.mathworks.com/moler/ncm/walker.m)

- Eine Suche bei der US Patent- und Warenzeichen-Behörde ergibt 1.197 US-Patente, die den Begriff „Singular Value Decomposition“ enthalten. Das älteste davon, "Ein fiberoptisches System zur Inspektion von Sandwich-Lötverbindungen in integrierten Schaltkreisen", stammt aus dem Jahr 1987. Daneben finden sich Titel wie „Kompression von Oberflächen-Lichtfeldern“, „Methode zur seismischen Überwachung“, „Semantische Suchen in Peer-to-Peer-Netzwerken“, „Biochemische Marker der Hirnfunktion“ und „Mit Diabetes leben“.

[www.uspto.gov/patft](http://www.uspto.gov/patft)

## QUELLEN

► **On the Early History of the Singular Value Decomposition**  
[locus.siam.org/SIREV/volume-35/art\\_1035134.html](http://locus.siam.org/SIREV/volume-35/art_1035134.html)

► **Cleve's Corner Collection**  
[www.mathworks.com/res/cleve](http://www.mathworks.com/res/cleve)